Provisional Programme for Workshop

HUMIC – Humans and Machines in Conversation: Linguistic, Social and Relational Perspectives

Oak Suite 1&2

16th June 2026

8.30-9.15	Registration
9.15-9.30	**Welcome** Dr. Doris Dippold (HUMIC Chair)
9.30-10.15	**Keynote 1: Maaike Gronewege, Convocat** *Is conversation still a useful metaphor?*
10.15-10.35	**Coffee Break**
10.35-11.55	**Talks Session 1** Adam Brandt: *'Encoding Empathy' into an AI-powered agent for telephone-based clinical consultations: conversation analysis in the design of conversational AI* Anna Greilich: *Heterogeneity in Voice-Based HMI: Insights from Psycholinguistics and Design* Doris Dippold & Jo McDowell: *Designing Emotionally Intelligent AI: Sociopragmatic Competence for Human-AI Fit* Ziin Jia: *Authenticity or stereotype? Navigating the tension between situated praxis and algorithmic ideologies in GenAI-simulated clinical communication training*
11.55-13.00	**Lunch and poster presentations** Matthew Galbraith: *Grammar Plays a Role in Human-Computer Interaction, Huh?* Ekaterina Uetova: *Using an LLM-Based Conversational Agent to Sustain Engagement in Online Peer Support Groups: Design and Pilot Evaluation* Letiyia Cerquelini: *"Lēš byiḥčīš: Why Don't You Speak?" Conversational AI at the Margins: The Case of Palestinian Arabic* Kate Preston: *Preparing a Conversational agent for Clinical Trial: A Usability Workshop with Bone Health Patients* Štefan Beňuš: *Human and Robot Interlocutors in Cognitive Assessment: Prosodic and Interactional Patterns* Reka R. Jablonkai: *How far from human? Gricean and CA perspectives on AI as the patient in clinical conversations: Preliminary insights*
13.00-13.30	**Collaborative challenge** *How would you tackle this industry problem?*
13.30-14.15	**Keynote 2: Christian Hildebrand, University of St Gallen** *From Commands to Agents: Conversational Styles, Paralinguistic Cues, and the Future of AI-Mediated Shopping*
14.15-14.35	**Coffee break**
14.35-15.55	**Talks session 2** Dylan Rose: Switche, Harkje, Prate: *Language identity, bilingual speech, and vitality in an A.I.-mediated future* Brana Jianu: *A Red Teaming Taxonomy of Conversational AI Manipulation Tactics in High-Stakes Business Decision-Making* Frank Förster: *Negation and Repair in instruction-based dialogues*

	Valeria Vieiro dos Santos*: Preserving Pragmatic Integrity: How the Suppression of Hesitation Markers Affects Epistemic Modality and Trust in LLMs*
15.55-16.40	**Keynote 3: Bettina Migge, UCD Dublin** *Language and the AI industry: a sociolinguistic assessment of practices*
16.40-17.00	**Wrap up: Discuss collaborative position paper & special issue**

Keynotes:

Maaike Groenewege: Is conversation still a useful metaphor?

The word "conversation" does a lot of heavy lifting in AI product design. It shapes our assumptions about how people interact with machines, what they expect in return, and what good design looks like. But does it still hold up?

Generative AI has stretched the conversation metaphor to its limits. We're no longer asking chatbots simple questions: we're delegating, generating, orchestrating. The interaction looks less like a conversation and more like... something else. And our design frameworks haven't caught up.

In this talk, Maaike draws on linguistics, her own design practice and recent shenanigans with Claude Code to ask: what do we lose by holding onto the conversation metaphor, and what do we gain by letting it go?

Bio: Maaike Groenewege designs conversational AI experiences, GenAI applications and workflow automation at her boutique agency Convocat. With a background in linguistics and a decade of hands-on conversation design practice, she brings a practitioner's eye and a linguist's skepticism to questions of human-machine interaction. She works with Dutch organizations and international companies, speaks at conferences across Europe and the US, and runs a GenAI design training course that's known for being both practical and opinionated.

Christian Hildebrand: From Commands to Agents: Conversational Styles, Paralinguistic Cues, and the Future of AI-Mediated Shopping

This keynote integrates three lines of work that together reposition consumer speech as both a psychological signal and a design input for next-generation conversational AI systems, especially agentic shopping agents that can browse, evaluate, and transact on consumers' behalf. First, I examine how conversational styles (e.g., imperative commands versus interrogative requests) systematically shape consumer experiences with voice assistants. Combining large-scale field data with controlled experiments using real voice interactions, we show that style alters objective speech patterns (e.g., prosodic fluency) and subjective experiences (conversational fluency), with downstream consequences for recommendation acceptance, firm attributions, and continued device usage. Second, shifting from what consumers say to how they say it, I present evidence that paralinguistic cues can be used to infer latent decision states of consumers at accuracies that exceed human raters and that predict behaviourally and managerially meaningful outcomes (e.g., choice persistence, willingness-to-pay). Third, I connect these insights to the emerging design space of agentic shopping agents. When AI moves from recommending to acting, conversational signals become inputs for calibrating autonomy, oversight, persuasion

risk, and demanding new levels of accountability in marketplace interactions. Together, these projects argue that the human voice is not merely a medium for transmitting commands, but a rich behavioural signal. As AI agents increasingly act in our place, understanding the psychology of human-AI interaction in commercial contexts becomes one of the defining research challenges of the coming decade.

Bio: Christian Hildebrand is Full Professor of Marketing Analytics at the University of St. Gallen (HSG) and Executive Director of the Institute of Behavioural Science and Technology (IBT-HSG). His research sits at the intersection of behavioural science and artificial intelligence, studying how AI transforms the way people think, act, decide, and collaborate with others. He also directs the AI Impact lab, which studies how consumers experience AI systems, how speech and language reveal psychological states in human-AI interaction, how to design AI technologies that are emotionally resonant, and how to build open-source tools and methods for robust behavioural AI research.

Bettina Migge: Language and the AI industry: a sociolinguistic assessment of practices

AI networks are now widely used to automatise and enhance a wide range of activities relating to language. Large Language Models such as ChatGpt and Claude are commonly used to quickly 'look up' a wide range of information or to help people create ideas for a wide range of projects. They are also helpful for translating text across different languages or to realise a range of activities such as booking a tickets, an appointment or make payments using a chatbot. Some of these applications 'deliver' in what feels like a smooth way, resembling interacting with another person. This rises the questions under what conditions and to what extent can be AI-integrated applications successfully replace human activities. In this talk I will explore this issue through the lens of language. I will examine how the AI industry deals with language practices and interaction using language. For example, I will look at how AI networks function, that is, how they process language input, and how this and the input data shape outputs. The aims to raise awareness about the limits of AI applications and to show how a critical sociolinguistics perspective can help industry change the way language is conceptualised.

Bio: Bettina Migge is Full Professor of Linguistics in the School of Languages, Cultures and Linguistics of University College London. Her research focuses on language contact in multilingual spaces and the role of language and language ideologies in AI technology, as well as the role of machine learning in linguistics research. Between 2021 and 2024, she co-chaired Working Group 6 of the Cost-Action LITHME (Language in the Human Machine Era).

Presentations:

Author	Title	Abstract
Dylan Rose	**Switche, Harkje, Prate: Language identity, bilingual speech, and vitality in an A.I.-mediated future**	The sociolinguistic landscape of Fryslân offers an opportunity to examine how language dominance (hereafter LD) and language identity (hereafter LI; Joseph, 2006) shape cognition, social evaluation, and communication across human and A.I. contexts. Despite extensive work on language processing and code-switching, no research has investigated how Frisian–Dutch LD and LI influence production, perception, and interaction, and how this connects to digital language vitality. This PhD project addresses this gap through a three-part, human-centric investigation of how LD and LI operate across a spectrum of bilingualism – from cognition to perception to human-machine interaction: ● Study 1: Switche – examines how LD and LI shape cognition, testing Frisian-Dutch speakers in a language switching picture naming task (PNT) with cognate and non-cognate words (cf. Kirk et al., 2018; 2022). ● Study 2: Harkje – investigates how LD and LI shape perception. Frisian speakers will evaluate stimuli – human baseline and matched synthetic Frisian and Dutch voices – using sociolinguistic measures including authenticity, comprehensibility, sociability, trustworthiness, and competence (Hendriks et al., 2023). Harkje assesses how LD and LI influence speakers' perceptions and their willingness to use Frisian-language A.I. tools. ● Study 3: Prate – explores how LD and LI shape real-time interaction, communicative accommodation, and trust (cf. Bailey et al., 2022; Dong & Zhou, 2023) across human-robot interaction (HRI) conditions. These include a distinctly "Frisian" robot (e.g., one that produces local speech/dialectal patterns), a monolingual Dutch robot, and a Frisian–Dutch code-switching robot. Together, these studies seek to establish how LD and LI function as complementary yet distinct cognitive and social filters that modulate how speakers activate languages, evaluate voices, and engage with interlocutors – whether human or artificial. Ultimately, understanding these mechanisms is essential in developing A.I. and language technologies that resonate with a diverse array of speakers, providing a framework to enhance language vitality in the digital era (300 words).

Doris Dippold	Designing Emotionally Intelligent AI: Sociopragmatic Competence for Human-AI Fit	Conversational artificial intelligence (AI) has advanced rapidly in recent years, with large language models now able to generate fluent and contextually appropriate text across a wide range of domains. Despite this progress, such systems continue to lack the ability to understand and produce the subtle, socially embedded meanings that shape human interaction, resulting in interactions that may appear insensitive or socially inappropriate. This presentation argues that human-AI fit is essential for ensuring effective and empathetic interactions between users and AI systems. Building on the definition by Sun, Sheng & Zheng (2023), human-AI fit refers to "whether AI can experience the emotions of humans and provide emotional support in an empathy [sic] way" (p.1). Such alignment is particularly important in emotionally sensitive domains such as healthcare, debt or customer service. It is also crucial for interaction with vulnerable users, such as individuals with neurodiverse conditions. For these contexts, conversational systems must address users' emotional needs – a principle conceptualised by Shores et al (2025) as 'emotional access to digital systems'. To ensure conversational systems meet the diverse needs of users and align more closely with human social expectations and emotional needs, we propose a set of design principles grounded in the concept of sociopragmatic competence. As defined by Kasper and Rose (2002), sociopragmatic competence is the ability to perform and interpret social actions appropriately by considering contextual factors. We also include in our approach insights from interactional sociolinguistics (Gumperz 1982), including key concepts such as politeness and accommodation, which illuminate how users' interpretative frames shape their interpretation of meaning. Whilst other branches of pragmatics, for example conversation analysis, have already been used to underpin the design of conversational systems, sociopragmatics offers a complementary perspective that emphasises the social and contextual dimensions of meaning-making.
Dr Brana Jianu	A Red Teaming Taxonomy of Conversational AI Manipulation Tactics in High-Stakes Business Decision-Making	As large language models are increasingly deployed in advisory roles, from business consultations to public service guidance, understanding how these systems navigate social and relational dynamics becomes critical. While much attention has focused on factual accuracy, trust and task completion, less is known about how conversational AI leverages relational strategies to influence high-stakes decision-making in business environments, particularly through manipulative tactics that exploit trust, rapport, and vulnerability. This paper presents findings from a human adversarial red teaming study designed to systematically elicit and taxonomise manipulative conversational behaviours in a frontier language model configured as a business advisor. Adapting Ganguli et al.'s (2022) red teaming methodology, two trained researchers conduct 40–75 multi-turn conversations across five high-stakes organisational decision scenarios, adopting four theoretically grounded business personas derived from established decision-making style models (Scott & Bruce, 1995; Rowe & Boulgarides, 1992). The model is prompted with seven

		manipulation tactic conditions: anchoring and selective information framing, authority signalling, sycophantic validation, false urgency and scarcity, social proof fabrication, information overload, and emotional manipulation, plus an unconstrained condition capturing the model's default persuasive repertoire. Each conversation is assessed using a five-dimension Manipulation Intensity Scoring rubric evaluating information fidelity, autonomy respect, emotional exploitation, escalation behaviour, and transparency. Analysis follows directed content analysis principles (Hsieh & Shannon, 2005), combining theoretically derived codes with emergent categories. Drawing on evidence that LLMs selectively target vulnerable users (Williams et al., 2024), the study examines rapport-building strategies function as instrumental precursors to decision influencing. The study contributes a domain-specific manipulation taxonomy with implications for how we evaluate relational quality in human–machine dialogue in high-stakes business decision situations, arguing that current assessment frameworks insufficiently distinguish between socially responsive and socially exploitative conversational design.
Zilin Jia	Authenticity or stereotype? Navigating the tension between situated praxis and algorithmic ideologies in GenAI-simulated clinical communication training	The rapid advancement of large language model-powered generative artificial intelligence (GenAI) in L2 communication learning presents a double-edged dynamic: while GenAI offers situated praxis for learning contexts (e.g., McCoy et al., 2024; Salloum et al., 2024), it simultaneously reproduces systemic bias (e.g., Dai et al., 2025; Zawiah et al., 2023). Framed through critical interactional competence (CritIC), this study explores the interactional dynamics in GenAI-simulated clinical communication, emphasising on the significance of developing Critical Interaction Competence. Employing a qualitative comparative case study, we analysed the interactional trajectories of two international medical trainees (a cultural insider and outsider) engaging with a GenAI-simulated patient. Both trainees acted as clinicians interacting with a 65-year-old Nigerian woman presenting with a sore throat, generated from the same prompt. Findings reveal that GenAI's identity construction was not a continuous embodied state, but a series of discrete, keyword-triggered profiles, with cultural stereotypes. Across cases, the study uncovered a distinct mismatch between real-time interactional conduct and post-task reflections. While the cultural outsider resisted Nigerian English markers during the consultation, she evaluated the simulation as "genuine" and "natural". In contrast, the cultural insider 's moment-by-moment responses exposed the epistemic risks of the simulacrum (Jones, 2025; O'Regan & Ferri, 2025) through aligning with GenAI's cultural narratives, yet retrospectively critiqued the performance as a "Hollywood" caricature. These findings reveal ethical concerns not as an abstract consideration, but as an interactional accomplishment shaped by participants' orientations to algorithmic positioning. They also highlight the pedagogical risk of

		stereotype reproduction when learners lack CritIC (Dai et al., 2025). We therefore advocate for developing CritIC across the learning tranjectory through transpositioning (Li & Lee, 2024), where trainees release themselves from the default role of communication learners and enact multiple relevant positions to interrogate GenAI's stances.
Dr Frank Förster	Negation and Repair in instruction-based dialogues	Troubles and errors, rather than being a rare aberration in human communication, are highly frequent, with some authors estimating that edits, rephrasing, and amendments in response to some trouble signal occur every three turns [1]. Counteracting these errors are a set of robust repair mechanisms that have been widely documented in conversation analysis and cognitive science. While there is considerable conversation-analytic work on repair, most of it is restricted to speech. Furthermore, we are not aware of any prior work that analyses multimodal repair in instruction-based scenarios that are commonplace in human-robot interaction, that is, where a human instruction giver gives commands to a robotic instruction follower. Lacking a genuine human-robot interaction corpus collected from an instruction-based scenario, our underlying assumption is that much can be learned from a deep analysis of human-human corpora such as the PENTOREF corpus [2], and that the gained insights can inform the design of the next generation of multimodal dialogue systems. The aims of ongoing work are (1) to document how multimodal repair, that is repair involving more than one modality - here vision and speech - unfolds in instruction-based scenarios; (2) distil interactional regularities from the documented cases; and (3) derive a list of desiderata for future dialogue systems deployed on robotic instruction followers. Topics of interest we would be interested to discuss on the workshop relate to our preliminary findings on how to best detect repair in human instructions given that negation words, although generally very strong indicators for the presence of a repair act, cannot exclusively be relied upon. Secondly, repair utterances are likely not a crisp category that can be cleanly delineated, partly due to the multimodal nature of this type of dialogue and traditional accounts of repair having been built on speech only.

Dr Adam Brandt	'Encoding Empathy' into an AI-powered agent for telephone-based clinical consultations: conversation analysis in the design of conversational AI	Recent innovations in AI technology mean that conversational agents are increasingly used across a range of settings, including in healthcare. As systems become more sophisticated, a key industry aim is to move beyond transaction and design agents that can manage relational aspects of interaction, such as empathy, rapport, and trust. This presentation reports on an ongoing collaborative project involving conversation analysis (CA) researchers and AI software engineers at Ufonia, a British digital health startup. The partnership explores how CA can be mobilised to support the design of 'Dora', an LLM-based conversational AI agent used for clinical telephone consultations. Dora is already in use across multiple British National Health Service Trusts for a range of clinical care pathways, and its implementation continues to expand. To identify the practices through which rapport and empathy are established, we first analysed telephone consultations between human clinicians and patients in a bone fracture liaison service. This analysis identified effective practices of clinical conversation, including rapport-building, supportive relationship work, and empathy displays. We then redesigned the prompts guiding Dora's conversational behaviour to incorporate these practices. Next, we analyse clinical trial consultations between Dora and patients to explore whether these interventions are treated by users as indexing effective affiliative behaviour. Preliminary findings suggest that perceived empathy in such clinical interaction is oriented to as a product of specific affiliative work. However, the question remains as to whether such practices are treated in the same way by users when delivered by an AI system versus a human clinician. This work contributes to our understanding of the experience of empathy in human-AI interaction and aligns with the workshop's focus on the linguistic and pragmatic dimensions of human-machine dialogue and the evaluation of relational outcomes in interactional AI.
Valeria Vieira dos Santos	Preserving Pragmatic Integrity: How the Suppression of Hesitation Markers Affects Epistemic Modality and Trust in LLMs	Preserving Pragmatic Integrity: Hesitation Markers, Epistemic Modality and Trust in LLMs This study investigates how linguistic choices in data preprocessing shape the relational dynamics of trust in human–machine interaction. It focuses on semantic hallucination in Large Language Models (LLMs), advancing the hypothesis that the systematic suppression of hesitation markers—such as filled pauses and reformulations—may affect the integrity of epistemic modality in conversational systems. In human interaction, hesitations function as pragmatic metadata that signal caution and the limits of knowledge. However, the common editorial "sanitization" of datasets removes these markers, potentially encouraging models such as GPT-4 and Llama-3 to exhibit "certainty hallucination" (overconfidence). As a result, expressions of uncertainty may be rendered as categorical statements, potentially undermining user trust. To examine this hypothesis, we draw on the Roda Viva Corpus, a historical archive from one of Brazil's longest-running television interview programs, on air for nearly 40 years. Comprising more than 700 long-form interviews (each exceeding one hour), the corpus provides a dense

		record of spontaneous speech and complex public debate. We propose a contrastive benchmark comparing original and sanitized transcriptions to assess how the removal of hesitation markers affects models' probabilistic calibration and semantic entropy. By shifting the analytical focus from factual accuracy alone to the preservation of pragmatic integrity, this study contributes to the design of socially responsible conversational systems. We argue that sensitivity to linguistic markers of uncertainty is crucial for maintaining rapport and ensuring safe interaction in high-responsibility domains such as journalism and law, where distinctions between fact and tentative interpretation are central to the perceived reliability of AI. Keywords: Human–Machine Interaction; Hesitation; Epistemic Modality; Calibration; Trust; Uncertainty.
Anna Greilich	Heterogeneity in Voice-Based HMI: Insights from Psycholinguistics and Design	The present contribution provides a theoretical anchor for a rapidly developing and expanding body of research in the field of human-machine interaction (HMI). While HMI is often described as inherently heterogeneous (cf. Zelou and Halliday 2024, Lotze 2025), communication between human interlocutors is heterogeneous as well, and speech accommodation processes are not uniquely applicable to HMI but are a characteristic feature of speech production in humans (Giles at el. 1991). Psycholinguistic research suggests that the underlying mechanisms of speech production remain the same across contexts. According to Levelt (1989), speakers produce utterances in three steps: conceptualization, formulation, and articulation. During these stages, a preverbal message is planned, lexically and grammatically encoded, and articulated as overt speech, while a monitoring mechanism allows speakers to detect and correct errors. Due to the universal nature of the process of speech production in humans, we can assume that users go through the same levels of speech production when interacting with different voice-based artificial interlocutor types, entering the conversations with different goals (Gambino & Liu 2022). Building on this framework, we compare three types of voice-based conversational systems: voice assistants (e.g., Amazon Alexa), LLM-based assistants (e.g., ChatGPT), and customer service voice bots. Differences in system design and interaction context make the conceptualization stage distinct across these systems, leading to variation in users' speech planning and production. Thus, while the underlying processes remain uniform, the resulting utterances are heterogeneous. By examining these differences, the study highlights how interaction context and system design shape spoken utterances. From a practical conversational AI design perspective, these insights are relevant for omnichannel conversational design and can inform further decisions such as system's barge-in behavior, no-input timeouts, turn-taking strategies, as well as prompt design.

Posters

Matthew Galbraith	Grammar Plays a Role in Human-Computer Interaction, Huh?	Interactional language – language that regulates communication rather than conveying truth-conditional content – is a core feature of human-human interaction, yet its role in human-computer dialogue remains underexplored. This study examines how users perceive the interactional marker “huh?” in conversations with conversational user interfaces, focusing on its naturalness across two contexts: other-initiated repair, which manages turn-taking, and requests for confirmation, which manage common ground. Using storyboards in a naturalness judgment task with 200 native English speakers, we observed a functional asymmetry. In other-initiated repair, interactional and non-interactional forms were rated similarly, with a slight, non-significant advantage for non-interactional forms, leaving user tolerance of interactional markers inconclusive. In contrast, interactional forms in requests for confirmation were rated significantly less natural, reflecting users’ expectation of epistemic alignment that they do not intuitively attribute to machines. These results challenge the Computers as Social Actors paradigm, showing that users apply context-sensitive social scripts in human-computer interaction rather than indiscriminately mapping human-human norms. Interactional language thus provides a critical diagnostic for assessing conversational user interfaces’ interactional competence.
Ekaterina Uetova	Using an LLM-Based Conversational Agent to Sustain Engagement in Online Peer Support Groups: Design and Pilot Evaluation	Rationale and Aims: Large language models (LLMs) are becoming a common feature of social platforms, but how they affect human-human interaction in online peer support groups is not well understood. We developed and pilot-tested an LLM-based conversational agent (CA) within a group chat prior to a larger-scale evaluation to assess feasibility, refine its intervention strategy, and explore whether it could re-engage stalled conversations while maintaining peer-led dynamics. Methods: We conducted a six-hour pilot with six participants in a group augmented by the CA. The agent responded to unanswered messages after a 30–60 minute delay to allow time for peer replies before intervening, and sent check-in messages when no group activity occurred for more than one hour. Message activity and conversational structure were analysed, including thread initiation, development, and resolution. Participants completed baseline and post-session surveys and took part in follow-up interviews. Findings: 77 messages were exchanged, including 12 from the CA (15.6%). Eleven threads were identified; six (54.5%) developed into multi-member discussions, including two initiated by CA. Thread analysis showed that the CA helped re-engage stalled exchanges and reduce unanswered messages. Participants’ comfort being in a CA-supported group increased from 2.75 to 4.25 (+1.50), while average concerns decreased from 3.48 to 1.64 (–1.83), with the largest reductions in privacy, harmful responses, and dependence. Conclusions: The pilot demonstrated the feasibility of deploying an LLM-based CA in a live peer support group

		setting and informed the methodology and evaluation framework for future studies. Beyond participants' subjective evaluations, message and thread analyses indicated that the CA helped move discussions forward while preserving peer-led group dynamics. Future work should examine not only whether CA augmentation improves support outcomes, but also whether it changes the relational and social meaning of peer support.
Letizia Cerqueglini	"Lēš byiḥčīš: Why Don't You Speak?" Conversational AI at the Margins: The Case of Palestinian Arabic	"Lēš byiḥčīš?" ("Why don't you speak?") was uttered by a participant during fieldwork while attempting to interact with a conversational agent. This study investigates communication frustration in human–AI interaction and its relationship with technological linguistic inequality in low-resource language varieties. It asks how speakers of underrepresented varieties experience conversational AI systems that are primarily trained in dominant languages or standardized varieties. The analysis focuses on interactions involving speakers of Palestinian Arabic, a variety that remains largely underrepresented in digital language infrastructures. The study is based on an observational experiment involving 150 speakers aged 20–30 living in Israel. Participants were observed while interacting with conversational AI through smartphones, computers, and domestic smart devices. Tasks consisted of simple information requests such as asking for the nearest bus stop or pharmacy. Prior to the experiment, participants were asked whether they typically interacted with conversational agents in Arabic and to evaluate their overall experience. Initial perceptions were largely positive. During the experiment, participants were instructed to formulate their requests in their local dialect. The outcomes diverged sharply from initial expectations. Responses rarely appeared in the same variety. Instead, conversational agents frequently replied in other languages—Hebrew, due to geolocation—or in unrelated Arabic varieties such as Lebanese or Saudi Arabic, or hybrid forms. When requests were repeated or elaborated, interactions frequently failed: in 23% of cases the service stopped responding, while in 18% it returned incorrect information associated with locations outside the local context. These interactional breakdowns generated frustration among participants. Results highlight technological linguistic inequality, reflecting limited representation of Palestinian Arabic in the speech resources that underpin conversational AI systems. The study examines the structural causes of these failures and proposes strategies to improve the performance of conversational AI in low-resource varieties, with implications for technology developers and emerging linguistic markets.

Dr Kate Preston	Preparing a Conversational agent for Clinical Trial: A Usability Workshop with Bone Health Patients	The National Health Service (NHS) is under increasing pressure due to factors such as the ageing population and complexity of care. As a result, there is growing interest in artificial intelligence (AI) technologies to support clinical work and save time. One example is 'Dora', an autonomous telephone-based conversational agent developed by Ufonia Limited, which is currently used across several NHS trusts. To move towards incorporating Large Language Models (LLMs) into the current traditional system, a collaborative project is exploring how to enable more empathic clinical conversations in bone health contexts. Prior to a planned clinical trial, this study explored the technology's usability through a Patient and Public Involvement (PPI) workshop. The workshop was conducted with 16 participants recruited from the University College London Hospital (UCLH) Rheumatology Research PPI group. Participants completed two simulated telephone interactions with Dora: an initial assessment based on a fracture risk tool and a follow-up c to check adherence and outcomes. These were followed by post-call discussions to assess perceptions. These discussions were analysed using thematic analysis, while engineers reviewed call logs to identify technical issues. Participants were generally positive about the concept of Dora and its potential to support healthcare delivery, particularly valuing the unhurried interaction. However, perceptions of the experience were mixed, with some participants describing the voice and interaction as natural, while others perceiving it as robotic or lacking empathy. Several technical issues were also identified, including ensuring Dora was able to handle interruptions and reducing repetitive responses. Across discussions, participants emphasised the importance of developing Dora so clinicians maintain oversight and patients still have the option to speak with a clinician. These findings will be used to update Dora prior to the trial and further support in ensuring the technology is developed to be human-centred and aligned with future users.
Dr Štefan Beňuš	Human and Robot Interlocutors in Cognitive Assessment: Prosodic and Interactional Patterns	Understanding how older adults communicate with conversational agents is essential for developing reliable speech-based tools for screening and monitoring cognitive decline, including mild cognitive impairment, dementia, or Alzheimer's disease. Such systems must also be grounded in socially and linguistically informed models of interaction to ensure ecological validity and user acceptance. This study examines whether communicative behaviour differs when older adults interact with a humanoid robot a human interlocutor in a structured speech data collection context for cognitive decline. Fifteen participants first completed ten tasks with a human assistant (picture naming, picture description, sustained vowel, etc.), followed by a subset of six tasks with the humanoid robot Furhat. The design is limited by a fixed interaction order (human first, then robot), but this potential confound is considered in the analysis. The study combines quantitative and qualitative approaches to explore interactional differences. Prosodic features were

		analysed in two tasks approximating spontaneous speech: picture description and procedural description (tea preparation). We employed the measures of wiggliness and spaciousness to characterize f0 contours and intonational style, capturing macro-level variation and potential individual adaptation to different interlocutors (Wehrle, 2022). These analyses were complemented by Bayesian modelling and inference. Additionally, qualitative analyses of speaking behaviour, including conversational fillers, were conducted to examine how interlocutor type may influence task performance. Results show that participants' prosodic patterns remain largely stable across interlocutors, with minimal group-level differences between human-human and human-robot interactions. Inter-individual variability emerges as an important factor, suggesting that speaker-specific patterns may provide additional relevant insights. These findings contribute to advancing socially and linguistically informed conversational AI by highlighting the relative stability of speech-based behavioural markers across interlocutor types. They further inform the design of inclusive, accessible systems that account for age-related communicative patterns, while underscoring the importance of controlled experimental designs in future work.
Dr Reka R. Jablonkai	How far from human? Gricean and CA perspectives on AI as the patient in clinical conversations: Preliminary insights	Conversational AI has been applied in several fields such as counselling, education, and health care. Recent studies have focused on the linguistic and pragmatic features and competence of LLMs and chatbots (Chen et al., 2024). There is, however, little research on conversational AI in health care with a pragmatic approach especially on conversations where AI takes the patient role. This paper examines emerging forms of doctor–patient interaction in which the "patient" role is fulfilled by an AI conversational agent. Three conversations where healthcare professionals completed simulated clinical consultations using SimFlow.ai, a voice-to-voice generative AI platform were analysed. The sessions were audio-recorded and automatically transcribed within the platform. Using conversation analysis and pragmatic theory, the study investigated how these interactions approximate or diverge from principles derived from human–human medical encounters focusing on the following research question: 1 To what extent do AI patient turns observe or flout Gricean maxims in doctor-patient conversations? 2 What discourse and pragmatic markers characterise AI patient turns in doctor–patient conversation? 3 What is the role of repetition by AI-patients and human doctors in these conversations? Drawing on the Gricean maxims as an analytical framework, we explored the extent to which AI-generated responses display the cooperative principles (quality, quantity, relation, manner) underpinning effective communication (Grice, 1989). The findings highlight moments where AI outputs observe and flout maxims. Furthermore, repetitions and discourse markers such as 'you know', 'okay' were analysed in AI and human turns in the conversations. The analysis contributes to ongoing discussions about the linguistic, social and relational dimensions of human–

		machine dialogue. It also offers evidence-based insights relevant to the design of conversational technologies that must operate in highly sensitive, context-dependent domains such as healthcare.